DOCUMENT RESUME

ED 456 386 CG 031 183

AUTHOR Fouladi, Rachel T.; McCarthy, Christopher J.; Moller, Naomi P.

TITLE Paper-and-Pencil or Online? Evaluating Coping and Attachment Measures.

PUB DATE 2001-08-00

NOTE 36p.; Paper presented at the Annual Meeting of the American Psychological Association (109th, San Francisco, CA, August 24-28, 2001).

PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.

DESCRIPTORS Affective Behavior; Attachment Behavior; College Students; *Computer Assisted Testing; Higher Education; *Internet; *Personality Measures; *Psychological Evaluation; *Psychological Testing

IDENTIFIERS Inventory of Parent and Peer Attachment; Negative Moods

ABSTRACT

In recent years many researchers, clinicians, and testing companies have shifted from paper-and-pencil to computer administrated tests of psychological measures. This paper explores the feasibility of using the Internet to collect data from the Inventory of Parent and Peer Attachment (IPPA), the Negative Mood Rating Scale (NMRS), and the Trait Meta-Mood Scale (TMMS) to assess psychological functioning. Data were collected from college students (N=164) in upper division elective courses at a large southwestern university. The online sample consisted of 234 subjects between the ages of 19 and 30. The psychometric, distributional, and factor analytic results evidenced some differences between the respondents but in general, the results showed adequate internal consistency and construct validity of the scale scores for both modes of assessment. Findings are in congruence with other research and suggest the viability of administering psychological questionnaires via computer technology and the Internet. The results open doors for researchers and counselors to use information obtained from online assessments to examine links between attachment and emotional functioning. (Contains 53 references and 5 tables.) (JDM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

Running Head: Paper-and-pencil or online

ED 456 386

Paper-and-pencil or online?

Evaluating coping and attachment measures

Rachel T. Fouladi

M.D. Anderson Cancer Research Center

Christopher J. McCarthy Naomi P. Moller

University of Texas at Austin

Poster presented at the Annual Convention of the American Psychological Association, San Francisco, August 24 – 28, 2001.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
C. M. CARTHY
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

CG031183

BEST COPY AVAILABLE

Abstract

This exploration was conducted to explore the viability of using the World Wide Web to collect data from three widely used measures of parental attachment and emotional adjustment. Data were collected from two comparable groups of participants and differences in response patterns on paper-and-pencil and World Wide Web versions of the measures at both the item level and scale score level, were documented. Importantly, however, the magnitude of the effects were in general very small. The basic similarity of the properties of the measures using paper-and-pencil and online internet modes of administration suggests both the viability of the Internet for assessing psychological phenomena and the importance of continued evaluation of technology mode effects.

Paper-and-pencil or online? Evaluating mode effects on measures of emotional functioning and attachment

In the five years since it became possible to post assessment inventories on the Internet (Musch & Reips, 2000), the attractiveness of the World Wide Web as a medium for collection of psychological information and research has led to a surge of studies and articles (Birnbaum, 2000). The advantages of collecting information this way are numerous and include: decreased experimenter demand (Buchanan, 2000) and social desirability effects (Joinson, 1999), reduced missing data (Stanton, 1998), avoidance of data entry errors since responses are entered directly into the database (Pasveer & Ellard, 1998), savings of money and time (Pasveer & Ellard, 1998) and possibly greater self-disclosure by participants (Davis, 1999). However, these potential advantages do not guarantee the generalizability of assessments conducted using the Internet (Pasveer & Ellard, 1998).

Potential problems associated with using the Internet for assessment include (a) the lack of a controlled environment that allows responding to measures at whatever time and in whatever setting suits the respondent, also allowing for repeat or mischievous responding (Buchanan, 2000), (b) important differences in the layout of questionnaire items depending on the respondents' browser software and settings (Baron & Siepmann, 2000), and (c) potential violations of privacy and issues of data security (Cho & LaRose, 1999). In addition, studies that have explored the comparability of computer (but non-Web) administrations with paper-and-pencil data collection have raised concerns about whether negative attitudes towards computers affected responses to computerized instruments (Scheulenberg & Yutrzenka, 1999; Tseng, Macleod & Wright, 1997). Several studies have also found subtle differences between data

collected in the two modalities (Miles & King, 1997; Potosky & Bobko,1997; Webster & Compeau, 1996; Schwartz, Mullis & Dunham, 1998).

As Buchanan (2000) points out, these problems suggest that equivalent reliability and validity cannot be assumed for internet-collected and "traditional" paper-and-pencil assessments. Thus far fewer studies have explored differences between Web and traditional survey responses (Krantz & Dalal, 2000) but those that have (e.g., Pasveer & Ellard, 1998; Stanton,1998) have found "remarkable congruence" (Krantz & Dalal, 2000, p. 35). This accords with the results of a number of studies comparing paper-and-pencil versus computer (not Web) administration of instruments. For example, equivalence has been found for instruments assessing numerous dimensions, including career attitudes and interests, emotional well-being, marital adjustment, and symptomatology (e.g., DiLalla, 1996; Donovan, Drasgow & Tahira, 2000; Hansen, Neuman, Haverkamp. & Lubinski, 1997; Parks, Meade & Johnson, 1985; Pinsoneault, 1996; Schmitz, Hartkamp, Brinschwitz, & Michalek, 1999).

Given the multiple potential differences between online and paper-and-pencil assessment, the degree of equivalence found thus far is both surprising and worthy of further investigation. This study used samples that are highly similar, so that any differences found between internet and paper-and-pencil results might be attributed to differences in the study environment and presentation of the instruments, rather than to differences in the study populations. Two related areas of psychological assessment seem very promising for such an investigation: (1) the development of instruments to measure parental attachment in adolescence and adulthood, and (2) the development of instruments to measure emotional regulation processes. These domains were chosen because they are currently in wide use by researchers and clinicians and have a demonstrated association to overall well-being.

With respect to the former, Armsden and Greenberg developed the Inventory of Parent and Peer Attachment (IPPA, Armsden & Greenberg, 1987) to measure attachment to both one's mother and father, and several studies have demonstrated the utility of this instrument with adolescent and adult populations (Brack, Gay, & Matheny, 1993; McCarthy, 1998; McCarthy, Moller & Fouladi., 2001). Bowlby (1988) suggested that attachment refers to the emotional bond experienced with another who is sensed as a source of security and who provides a secure base anchoring exploration. A secure attachment therefore is hypothesized to contribute to autonomy and competence. Attachment theory has also been labeled a theory of affect regulation (Feeney & Noller, 1996). Accordingly, attachment is thought to be an important aspect of personality, since individual differences in attachment reflect rules and strategies that children learn about handling emotions that can persist across the lifespan. For this reason, continuing attachment to parents in adolescence and adulthood has been hypothesized to have an ongoing impact on emotional functioning (Braver, Burnberry, Green, & Rawson, 1992; Gilbert, 1992). However, attachment theory, particularly as it applies to adolescents and early adults, is not yet a complete model but rather a set of propositions in need of clarification and empirical verification (Kenny & Rice, 1995). To this end, McCarthy et al. (2001) demonstrated a clear relationship between attachment and emotional functioning.

Beyond the context of attachment theory, emotions and emotion-related processes are central to many, or most, theories of personality, but systematic research is needed in this area as well (Heesacker & Bradley, 1997). In part to address this gap in the research, Catanzaro and Mearns (1990) developed the Negative Mood Regulation Scale (NMRS) to measure beliefs about one's ability to alleviate negative moods. Similarly, Salovey, Mayer, Goldman, Turvey, and Palfai (1995) examined a construct they labeled "meta mood," which is also addressed in

their work on emotional intelligence (Mayer, DiPaolo, & Salovey, 1990). Salovey et al. (1995) developed the Trait Meta-Mood Scale (TMMS) to measure the more enduring aspects of this construct of the reflective experience of mood. Although these two bodies of research overlap, Catanzaro and Mearns' (1990) work is embedded in the tradition of social learning theory, in which theorists view generalized expectancies for problem solving as important determinants of an individual's behavior in a given situation. By contrast, Salovey et al. (1995) conceptualized emotional intelligence as a component of a more generalized set of human intelligences, which includes other abilities such as linguistic, musical, and bodily-kinesthetic intelligence.

As researchers continue to assess the relationship between emotional functioning and attachment and as the Internet is increasingly seen as a valid mode for conducting assessments and research, more studies of emotional and psychological functioning will likely be conducted using computer technologies. This study therefore sought to explore the comparability of paper-and-pencil versus Internet versions of three important and widely used instruments to assess psychological adjustment: the IPPA, TMMS and the NMRS.

Method

Participants

All participants were college students taking upper division elective courses at a large southwestern university.

The paper-and-pencil sample consisted of 164 students, 59% of whom were female, 41% male, and 1% providing no sex identification. The average age of the study participants was 20.71 (SD=3.98, range 18 to 45). The participants were 70% European American, 13% Latino(a), 9% Asian American, and 2% African American; 6% identified themselves as belonging to other ethnic groups.

The online sample consisted of 234 subjects, 61% of whom were female, 38% male, and 1% provided no sex identification. The mean age of the sample was 21.69 (SD = 1.52, range 19 to 30). Sixty-one percent of respondents self-reported their ethnicity as Anglo, 14% Asian, 4% African-American, 13% Hispanic, 3% Biracial, 2% Multiracial, 2% Other, and 1% provided no ethnic identification.

Procedures

Participants were recruited from undergraduate educational psychology classes over several semesters. For the online study, participants were provided with a World Wide Web address where they completed various attachment and emotional function measures as well as a demographic survey. For the paper-and-pencil study, participants were provided with a packet of instruments that they completed under supervision of an experimenter.

Instrumentation

Inventory of Parental and Peer Attachment (IPPA). This inventory assesses affective and cognitive dimensions of the current attachment of college students and adolescents (Armsden & Greenberg, 1987; 1989). There are 25 5-point Likert-type items on each of three scales measuring attachment to the mother, father, and peers. However, the present studies included only the two 25-item instruments of parent attachment (IPPA-M and IPPA-F). The IPPA-M measures current respondent attachment to mother; the IPPA-F measures current respondent attachment to father.

While an earlier version of the IPPA assessed attachment to parents as a single construct (Armsden & Greenberg, 1987), the authors later revised the scale to assess attachment to mother and father separately (Armsden & Greenberg, 1989). Participants are asked the same questions, once for mother and once for father. Examples of items include: "I get upset a lot more than my

mother/father knows about", "When we discuss things, my mother/ father cares about my point of view", "My mother/father trusts my judgment. This revised version of the instrument has been used in several studies of late adolescent attachment (Brack et al., 1993; McCarthy, Brack, Brack, Liu, & Carlson, 1998). Armsden and Greenberg (1987) reported internal consistency (Cronbach's alpha) scores that ranged from .86 to .91 and test-retest reliability scores over a three-week period of .93 for their overall parental attachment scale; internal consistency estimates for scores from the separate mother and father scales has been reported at .89 and .88 respectively (Papini, Roggman & Anderson, 1991). In a recent study with a sample of college students having both mothers and fathers, Cronbach's alpha internal consistency estimates were .93 for attachment to mother (IPPA-M) and .95 for attachment to father (IPPA-F), and the observed correlations between the scores from the two instruments was .40 (McCarthy et al., 2001). Because of Armsden and Greenberg's arguments for the separate assessment of attachment to each parental figure, IPPA-M and IPPA-F are treated as separate single factor instruments in this study. In this study, scores on each of the 25-item instruments are established from the summing of appropriately coded items; scale scores have a possible range of 25 to 125 with higher values representing stronger attachment.

Negative Mood Rating Scale (NMRS). The NMRS is a 30-item inventory with three scales that measure perceived ability to reduce negative mood (Kirsch, Mearns, & Catanzaro, 1990). The Cognitive scale assesses confidence in using cognitive strategies to reduce negative mood; the Behavior scale measures expectancies about using overt behaviors to change negative emotions; and the General scale assesses generalized beliefs that one can alter one's mood. Items on each scale are 5-point Likert-type items. Examples of items from the three scales are: "I can forget about what's upsetting me pretty easily" (Cognitive), "I can usually find a way to

cheer myself up" (General), "Doing something nice for someone else will cheer me up" (Behavior). Internal consistency estimates reported from six separate samples found Cronbach alphas ranging from .86 to .92 (Catanzaro & Mearns, 1990). In the current study, scores on each of the NMRS dimensions are reported as the mean response on appropriately coded items; scale scores have a possible range of 1 to 5 with higher values representing stronger beliefs about one's own ability to reduce negative mood.

A number of studies have documented the utility of the NMRS. Kirsch et al. (1990) found that NMRS scores predicted dysphoria and somatic symptoms in a sample of college students. Catanzaro and Greenwood (1994) also demonstrated that NMRS scores were positively related to active coping behaviors and negatively related to avoidant coping and stress symptoms.

Trait Meta-Mood Scale (TMMS). The TMMS is a 30-item instrument designed to measure awareness of mood and mood regulation strategies (Salovey, Mayer, Goldman, Turvey, & Palfai, 1995). It includes three scales that assess long-term aspects of the reflective experience of mood: tendency to attend to mood (Attention), the ability to discriminate different feelings (Clarity) and the ability to regulate mood (Repair). Examples of items are: "I try to think good thoughts no matter how badly I feel" (Repair), "People would be better off if they felt less and thought more" (Attention), "Sometimes I can't tell what my feelings are" (Clarity). Items on each scale are 5-point Likert-type items. Cronbach alphas for scores on these scales range from .82 to .87 (Salovey et al., 1995). Mayer and Stevens (1994) found evidence that the TMMS scales were related to criterion variables such as coping behaviors and personality functioning. In the current study, scores on each of the TMMS dimensions are reported as the mean response on appropriately coded items; scale scores have a possible range of 1 to 5 with higher values

representing greater awareness of one's own mood and mood regulation strategies.

Analysis and Results

Item Analyses

The psychometric characteristics of each of the scales was examined with the data collected via the two modes of administration (paper-and-pencil vs. online internet). These analyses were based on the study participants who responded to every item of a given instrument. The estimated properties (e.g., item means, item variances, inter-item correlations, item-total correlations, coefficient alpha, maximum coefficient alpha if any single item is deleted) of each measure show the substantial similarity across the two modes of delivery. The number of study participants responding to every item on each scale and summary psychometric properties are reported in Table 1.

To assess the impact of administration mode on item means and variances a series of multivariate analysis of variance and Box's M F-tests of homogeneity of covariance matrices were conducted for each set of scale items. The multivariate tests showed statistically significant mean differences on the set of items for the IPPA-M scale, the NMRS Cognitive and Behavior scales, and all three TMMS scales (Attention, Clarity, and Repair; p's $<.05$ for ε^2 ranging from .051 to .142). No statistically significant multivariate mean differences were indicated for the IPPA-F scale ($p=.444$) or the NMRS General scale ($p=.450$). Statistically significant heterogeneity of covariance matrices was indicated for all of the scales (p's$<.01$). Table 2 summarizes the results of these analyses.

A series of follow-up univariate analyses were conducted for items of the scales where multivariate statistics showed statistically significant results. These univariate analyses included independent samples t-tests and Levene's tests of homogeneity of variances. These univariate

analyses showed statistically significant mean differences (p's <.05) for the following number of items on each of the scales: 7 of 25 IPPA-M items (Maximum ε^2=.048), 1 of 10 NMRS Cognitive items (ε^2=.030), 2 of 10 NMRS Behavior items (Maximum ε^2=.021), 2 of 13 TMMS Attention items (Maximum ε^2=.016), 2 of 11 TMMS Clarity items (Maximum ε^2=.016), and 2 of 6 TMMS Repair items (Maximum ε^2=.010). Statistically significant variance differences were found for: 9 of the 25 IPPA-M items, 1 of the 25 IPPA-F items, 1 of the 10 NMRS Cognitive items, 3 of the 10 NMRS Behavior items, 3 of the 10 NMRS General items, 8 of 13 TMMS-Attention items, 5 of the 11 TMMS-Clarity items, and 2 of the 6 TMMS Repair items (all p's <.05). Thus, across the 110 items, statistically significant mean and variance differences were shown on 16 and 32 items respectively. The magnitude of these differences however were small.

Further multivariate analyses were conducted to assess whether there were any additional differences on the items of each instrument (IPPA-M, IPPA-F, NMRS, and TMMS) as a function of sex and ethnic identification. Four multivariate general linear model analyses were conducted with sex (male vs. female), ethnic identification (European American/Anglo vs. Other), as well as mode of administration (paper-and-pencil vs. online) parameterized as between-subject factors; included in the model were two-way interactions among the factors. Table 3 summarizes the results of the multivariate general linear model analyses. In these analyses, the only statistically significant effects identified were on the IPPA-M scale (sex , p=.044, ε^2=114; mode of administration, p=.003, ε^2=.142), NMRS (sex , p=.003, ε^2=.155), and TMMS (sex, p=.008, ε^2=.149). That there were no statistically significant interactions of sex or ethnic identification with mode of administration is noteworthy: the implication is that sex and ethnic effects were uniformly observed across the two modes of administration.

Composite Score Analyses

The multivariate comparability of the composite scores derived from the measures in the paper-and-pencil and online studies was examined using tests of homogeneity of mean vectors and covariance matrices. These analyses were conducted on those individuals who provided responses to at least 90% of the items underlying each of the measures (N=333). Scoring was conducted under the assumption of a consistent response pattern to the items underlying each scale. MANOVA results showed statistically significant differences between modes of administration on the set of measures for the two data sets (Wilks' L=.690, $F(8, 324)$=18.23, p<.001, ε^2=.310). Box's M F-test indicated statistically significant differences between the covariance matrices for the corresponding composite scores from two data sets (Box's M=188.15, $F(36, 349433.3)$=5.09, p<.001). Univariate comparisons showed statistically significant mean differences only on two of the TMMS measures: Attention and Clarity (p's<.001 for ε^2=.080 and .099 respectively: $t(385)$=-5.78, p<.001, M_{Paper}=3.54 vs. M_{Online}=3.85 on Attention; and $t(392)$=6.58, p<.001, M_{Paper}=3.83 vs. M_{Online}=3.46 on Clarity). Tests of homogeneity of variances showed statistically significant differences only on TMMS Repair ($F(1, 331)$=11.24, p=.001, SD_{Paper}=.56 vs. SD_{Online}=.73). No other univariate effects showed statistical significance.

Kolmogorov-Smirnov tests were conducted to assess the univariate normality of the data for each of the measures. Statistically significant departures from normality were exhibited in the case of the paper-and-pencil study for IPPA-M (p=.004; skew = −1.13, SE = .19; kurtosis = 1.08, SE = .38), and in the case of the online study for IPPA-F (p = .021; skew = −.83, SE = .17; kurtosis = .14, SE = .34), IPPA-M (p =.009; skew = −1.18, SE =.16; kurtosis = 1.43, SE= .33), NMRS Behavior (p =.009; skew = −.51, SE = .16; kurtosis = .47, SE = .32), and NMRS General

(p =.001; skew = –.83, SE = .16; kurtosis = .94, SE = .32). Because of the level of non-normality on some of the measures, but despite the known robustness of independent samples t-tests to modest levels of non-normality, a series of nonparametric tests (Mann-Whitney U, Kruskal-Wallis chi-square, and median chi-square tests) were further conducted to compare the distributions of each measure in the paper-and-pencil study with those in the online study. The results of the three non-parametric tests showed no statistically significant differences on the IPPA-F, the three NMRS measures, and TMMS Repair (p's>.05), and statistically significant differences on TMMS Attention and Clarity (p's<.001). In the case of IPPA-M, the results of the Mann-Whitney U and Kruskal-Wallis tests yielded p=.109, and the median chi-square test p = .019.

Of the eight composite scores examined only the TMMS Attention and Clarity scale scores consistently showed statistically significant and moderate differences in distributional parameter estimates. Further multivariate analyses were conducted to assess whether there were any additional differences on the set of scale scores as a function of sex and ethnic identification. A three-way multivariate analysis of variance was conducted with sex (male vs. female), ethnic identification (European American/Anglo vs. Other), as well as mode of administration (paper-and-pencil vs. online) parameterized as between-subject factors. Because the three-way interaction effect was not statistically significant, a model with only main effects and two-way interactions was examined. This analysis showed that though there were statistically significant main effects for sex (Wilks' L =.946, F(8, 316)=2.25, p=.024, ε^2=.054) and mode of administration (Wilks' L =.857, F(8, 316)=6.57, p<.001, ε^2=.143), and there was a statistically significant interaction of sex with ethnic identification (Wilks' L =.690, F(8, 316)=4.24, p<.001, ε^2=.097), there were no other statistically significant effects (all p's >.05). Of particular note is

that there were no statistically significant interactions of sex or ethnic identification with mode of administration. Assessment of the homogeneity of covariance matrices using Box's M F-test showed statistically significant differences in the covariance matrices ($F(216, 7266.7)=1.60$, $p<.001$); however univariate analyses comparing the variances did not yield any statistically significant effects.

Confirmatory Factor Analyses

Single factor models (IPPA-M and IPPA-F) and three-correlated factor models (NMRS and TMMS) were parameterized in confirmatory factor analyses (CFA). Multi-sample confirmatory factor analyses assess the homogeneity of each instrument structure across the two modes of instrument administration; further comparisons as a function of sex and ethnic identification were not conducted due to sample size constraints. Analyses were based on the study participants who responded to every item of every instrument ($N=251$). Because of the size of the multi-sample confirmatory factor analysis models, the resultant sample size was not sufficient for use of asymptotically distribution free generalized least squares covariance structure analysis procedures (Browne, 1984). Furthermore, because none of the current distributions of any of the structural equation modeling programs (EQS, AMOS, Mplus, and Statistica included) incorporate a Satorra-Bentler adjustment for the multi-sample fit statistics (Bentler, 1995; Arbuckle & Wothke, 1999) and there is no published method for a multi-sample Satorra-Bentler scaling, the following multi-sample results are based on unscaled maximum likelihood (ML) statistics (thereby providing an underestimate of model-data correspondence of the model to the data that would be indicated with Satorra-Bentler scaling). Reported multi-sample RMSEA results are appropriately scaled multi-sample RMSEA statistics (Note: many structural equation modeling programs implement a number of multi-sample goodness of fit

statistics, however the widely used RMSEA estimates provided by even the most recent versions of some of these statistical programs are off by a factor of the square root of the number of samples under comparison (c.f., Steiger, 1995, 1998; Muthén & Muthén, 1998) Table 4 provides summary CFA results.

Multi-sample analyses for each instrument yielded similar estimates in the unconstrained and constrained models. The unscaled multi-sample maximum likelihood results and corresponding RMSEA statistics indicate the imperfect fit of the theoretical models to the data (though, with a Satorra-Bentler modification the magnitude of the underlying chi-square statistic, RMSEA and confidence interval (CI) boundaries would be lower). Importantly, however with regard to the focus of the present paper (i.e., the comparison of paper-and-pencil and online mode effects), these unscaled results still show that there were no statistically significant differences between the paper-and-pencil and online models as indicated by the similar RMSEA values and confidence intervals and the results of the unscaled chi-square difference tests comparing the fit of the constrained with the unconstrained models (p's>.05). Thus, no mode effects were seen in the confirmatory factor analyses.

In the case of the comparison of the unconstrained models (MU) with the first set of constrained models specified with equality constraints on factor loadings and intercorrelations (MC1), the unscaled chi-square difference tests yielded no statistically significant difference in model-data fit on any of the instruments (D1 p's >.05). The imposition of further equality constraints on unique and factor variances in the second set of constrained models (MC2) also yielded no statistically significant differences in model-data fit (D2 p's >.05). Application of the RDR proposed by Browne and DuToit (1992) which is the equivalent of an RMSEA for the chi-square difference test yielded corresponding RDR point estimates ranging from .000 to .081 and

from .000 to .089 for the first and second set of chi-square difference tests respectively. Corresponding 90%CIs using the noncentrality interval estimation method described in Steiger and Fouladi (1997) were obtained. The analyses showed that the models were appropriately constrained, and that there was very good correspondence between the factor solutions obtained from the paper-and-pencil study with those obtained from the online study, insofar as there were no meaningful changes in model-data fit. That these results obtain even without a Satorra-Bentler rescaling suggests the robustness of the conclusions of the comparability of the measurement models for the measures in the two data sets in this study.

Because the results of the multi-sample comparisons suggest the general comparability of the factor solutions and the results of the mean and variance comparisons suggest only small differences between the means and variances of the data from the paper-and-pencil and online administrations of the measures, confirmatory factor analyses were conducted for the pooled data sets. The unscaled and Satorra-Bentler scaled maximum likelihood RMSEA and 90%CI results for these overall analyses were obtained; further scaling to address the sample size to model size ratio using Bartlett modifications of the Satorra-Bentler scaled maximum likelihood statistic (Fouladi, 1999) yielded decreases in the RMSEA point estimates and confidence interval limits of no more than .01. In total, these factor analytic results show that the theoretical factor models exhibit some correspondence to the data, that sample size was adequate to conduct the confirmatory factor analysis with relative precision; however, according to the criteria commonly used in confirmatory factor analysis, the adequacy of the theoretical models underlying these measures are in question, insofar as there clearly is good fit but certainly not "perfect" or "very close" fit (Steiger, 1989). Similar results obtained for each sample analyzed separately – with

the exception of TMMS online data that under Satorra-Bentler scaling showed excellent correspondence to the theoretical model.

Summary of Results

The psychometric, distributional, and factor analytic results evidenced some differences between the responses of study participants in the paper-and-pencil and online studies. However, overall the mode effects when present were small. In general the results show adequate internal consistency and construct validity of the scale scores, for both modes of assessment. Because of the overall comparability of the scale scores across the two modes of assessment, table 5 provides an overall summary of the descriptive statistics for the scale scores pooled across the two data sets. As such, the table provides an overview of the distributional characteristics in the current sample. In particular, the pattern of correlational results provide evidence of the convergent and divergent validity of the scales for this sample. Of particular note is that the attachment scale scores are more strongly intercorrelated than they are correlated with the NMRS and TMMS scales. Similarly, the NMRS scales scores are more strongly intercorrelated that they are correlated with the IPPA or TMMS scales. The pattern of intercorrelations of the TMMS scale scores are somewhat different however, with some of the scales showing stronger intercorrelations with NMRS scales than with other TMMS scales. Nonetheless, the intercorrelations among the emotional function scales (TMMS and NMRS) are stronger than the correlations with the attachment scales. Thus, though intercorrelated the attachment measures and emotional functioning measures are appropriately considered separately.

Discussion

In recent years many researchers, clinicians, and testing companies have shifted from paper-and-pencil to computer administration of psychological measures to assess psychological

function. The advent of Internet technology and the ability to post questionnaires on the World Wide Web further extends opportunities for assessing and researching psychological function. The possibilities of these new technologies have been exploited with computerized adaptive testing (Wainer, 2000), assessments using handheld devices (Stone & Shiffman, 1994), and Internet assessments (Birnbaum, 2000). However, as evidenced by the differences found between computer and paper-and-pencil administrations of various personality instruments (Miles & King, 1997; Potosky & Bobko,1997; Webster & Compeau, 1996; Schwartz et al., 1998), the comparability of data collected via traditional means and online cannot be assumed (Baron & Siepmann, 2000; Davis, 1999).

Using data on emotional functioning and parental attachment collected from two comparable groups of study participants, this study documented the occurrence of statistically significant mode effects between paper-and-pencil and Internet administrations at both the item and scale level. However, the magnitude of the corresponding effects on the means and variances of the items and scale scores was in general very small. The psychometric properties of the instruments were overall highly similar, and confirmatory factor analyses assessing the homogeneity of the factor patterns, intercorrelations, and variances indicated the congruence of the structure of each of the instruments across the two modes of administration, despite the presence of imperfect model-data fit. In sum, the general similarity of the properties of the measures using paper-and-pencil and Internet modes of administration is notable.

The findings in this study of substantial similarities and subtle differences between response patterns to measures of emotional functioning and attachment across two modes of questionnaire administration are consistent with findings by researchers investigating mode effects on other aspects of psychological function. These results also suggest the importance of

the continued evaluation of mode effects and the refinement of such measures. While demonstrating the viability of using different modes of administration to collect data, findings of small mode effects and imperfect model fit provide empirical support for the importance of continuing evaluation of commonly used assessment measures n general and attachment and emotional functioning in particular.

Several cautions need to be observed with regard to these findings. First, some of these analyses were limited by missing data. Second, because participants were all high school graduates, and enrolled in undergraduate courses, the homogeneity of the sample, with respect to ethnicity, age, and education level, limits the generalizability of the findings to other populations.

Despite the limitations, the results of this study may represent an important contribution. As Krantz and Dalal (2000) point out, few studies to date have explored differences between Web and traditional survey responses. Despite general findings of mode effects, those studies that have investigated the comparability of response patterns to paper-and-pencil and Web administered questionnaires have found few differences between response patterns (Pasveer & Ellard, 1998; Stanton, 1998). Importantly, the results of this paper and the studies of Pasveer and Ellard (1998) and Stanton (1998) are consistent with a number of studies establishing the comparability of paper-and-pencil with non-Web computer administration of several psychological instruments. These findings therefore suggest the viability of administering psychological questionnaires via computer technologies and the Web, and opens the door for researchers and clinicians to use information from the IPPA, TMMS, and NMRS that has been collected online to examine links between attachment and emotional functioning.

References

Arbuckle, J.L., & Wothke, W. (1999). AMOS 4.0 user's guide. Chicago, Illinois: SmallWaters Corporation.

Armsden, G., & Greenberg, M. (1987). The inventory of parent and peer attachment: individual differences and their relationship to psychological well-being in adolescence. Journal of Youth and Adolescence, 16, 427-454.

Armsden, G.C., & Greenberg, M.T. (1989). The Inventory of Parent and Peer Attachment: Preliminary Test Manual. (Available from G. C. Armsden, Department of Community Health Care Systems, University of Washington, Seattle, WA, 98195).

Baron, J., & Siepmann, M. (2000). Techniques for creating and using Web questionnaires in research and teaching. In M.H. Birnbaum (Ed.) Psychological experiments on the internet, Academic Press: San Diego, CA, pp. 235 –266.

Bentler, P.M. (1995). EQS: Structural equations program manual. Encino, California: Multivariate Software Inc.

Birnbaum, M.H. (2000). Psychological experiments on the internet. San Diego, CA: Academic Press.

Bowlby, J. (1988). Developmental psychiatry comes of age. The American Journal of Psychiatry, 145, 1-10.

Brack, G., Gay, M.F., & Matheny, K. B. (1993). Relationships between attachment and coping resources among late adolescents. Journal of College Student Development, 34, 212 - 215.

Braver, M., Burnberry, J., Green, K., & Rawson, R. (1992). Child abuse and current psychological functioning in a university counseling center population. Journal of Counseling Psychology, 39, 252-257.

Browne, M.W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. British Journal of Mathematical and Statistical Psychology, 37, 62-3.

Browne, M.W., & DuToit, S. (1992). Automated fitting of nonstandard models. Multivariate Behavioral Research, 27, 269-300.

Buchanan, T. (2000). Potential of the Internet for personality research. In M. H. Birnbaum (Ed.) Psychological Experiments on the Internet, Academic Press: San Diego, CA, pp121 – 140.

Catanzaro, S. J. & Greenwood, G. (1994). Expectancies for negative mood regulation, coping and dysphoria among college students. Journal of Counseling Psychology, 41 (1), 34-44.

Catanzaro, S. J., & Mearns, J. (1990). Measuring generalized expectancies for negative mood regulation: Initial scale development and implications. Journal of Personality Assessment, 54, 546-565.

Cho, H., & LaRose, R. (1999). Privacy Issues in Internet Surveys. Social Science Computer Review, 17, 421-434.

Davis, R.N. (1999). Web-based administration of a personality questionnaire: Comparison with traditional methods. Behavior Research Methods, Instruments and Computers, 31, 572-577.

DiLalla, D.L. (1996). Computerized administration of the Multidimensional Personality Questionnaire. Assessment, 3, 365-374.

Donovan, M.A., Drasgow, F., & Probst, T.M. (2000). Does computerizing paper-and-pencil job attitude scales make a difference? New IRT analyses offer insight. Journal of Applied Psychology, 85, 305-313.

Feeney, J., & Noller, P. (1996). Adult attachment. Thousand Oaks, CA: Sage.

Fouladi, R.T. (1999). Model fit in covariance structure analysis under small sample conditions - Modified maximum likelihood and asymptotically distribution free generalized least squares procedures. Presented at American Educational Research Association Annual Meeting, Montreal, April 19-23, Quebec, Canada.

Gilbert, S.P. (1992). Ethical issues in the treatment of severe psychopathology in university and college counseling centers. Journal of Counseling and Development, 70, 695-699.

Hansen, J.C., Neuman, J.L., Haverkamp, B.E., & Lubinski, B.R. (1997). Comparison of user reaction to two methods of Strong Interest Inventory administration and report feedback. Measure and Evaluation in Counseling and Development, 30, 115-127.

Heesacker, M., & Bradley, M.M. (1997). Beyond feelings: Psychotherapy and emotion. The Counseling Psychologist, 25, 201 - 219.

Joinson, A. (1999) Social desirability, anonymity and Internet-based questionnaires. Behavior Research Methods, Instruments and Computers, 31, 433-438.

Kenny, M.E., & Rice, K.G. (1995). Attachment to parents and adjustment in late adolescent college students: Current status, applications, and future considerations. The Counseling Psychologist, 23, 433-456.

Kirsch, I., Mearns, J., & Catanzaro, S. J. (1990). Mood-regulation expectancies as determinants of dysphoria in college students. Journal of Counseling Psychology, 37, 306-312.

Krantz, J.H., & Dalal, R. (2000). Validity of Web-based psychological research. In M. H. Birnbaum (Ed.) Psychological Experiments on the Internet, (pp. 35 – 60). Academic Press: San Diego, CA.

Mayer, J. D. & Stevens, A. A. (1994) An emerging understanding of the reflective (meta) experience of mood. Journal of Research in Personality, 28, 351-373.

Mayer, J. D., DiPaolo, M., & Salovey, P. (1990). Perceiving affective content in ambiguous visual stimuli: A component of emotional intelligence. Journal of Personality Assessment, 54(3&4), 772 - 781.

McCarthy, C.J. (1998). Relationship of parental attachment and gender to beliefs about mood. Journal of College Student Psychotherapy, 12, 39-59.

McCarthy, C. J., Brack, G., Brack, C. J., Liu, H. L., & Carlson, M. H. (1998). Relationship of college students' current attachment to appraisals of parental conflict. Journal of College Counseling, 1, 135-153.

McCarthy, C.J., Moller, N.P., & Fouladi, R.T. (2001) Continued attachment to parents: It's relationship to affect regulation and perceived stress among college students. Measurement and Evaluation in Counseling and Development, 33, 198-213.

Miles, E.W., & King, W.C. (1997). Gender and administration mode effects when pencil-and-paper personality tests are computerized. Educational and Psychological Measurement, 58, 68-76.

Musch, J., & Reips, U. (2000). A brief history of Web experimenting. In M. H. Birnbaum (Ed.) Psychological Experiments on the Internet, Academic Press: San Diego, CA, pp61-87.

Muthén, L.K., & Muthén, B.O. (1998). Mplus: The comprehensive modeling program for applied researchers user's guide. Los Angeles, California: Muthén & Muthén.

Papini, D. R., Roggman, R. L. & Anderson, J. (1991). Early-adolescent perceptions of attachment to mother and father: A test of the emotional-distancing and buffering hypothesis. Journal of Early Adolescence, 11, 258-275.

Pasveer, K.A., & Ellard, J. H. (1998). The making of a personality inventory: Help from the WWW. Behavior Research Methods, Instruments and Computers, 30, 309-313.

Parks, B.T., Mead, D.E., & Johnson, B.L. (1985). Validation of a computer administered Marital Adjustment Test. Journal of Marital and Family Therapy, 11, 207-210.

Pinsoneault, T.B. (1996). Equivalency of computer-assisted and paper-and-pencil administered versions of the Minnesota Multiphasic Personality Inventory-2. Computers in Human Behavior, 12, 291-300.

Potosky, D. & Bobko, P. (1997). Computer versus paper-and-pencil administration mode and response distortion in noncognitive selection tests. Journal of Applied Psychology, 82, 293-299.

Salovey, P., Mayer, J. D., Goldman, S. L., Turvey, C., & Palfai, T. P. (1995). Emotional attention, clarity, and repair: Exploring emotional intelligence using the trait meta-mood scale. In J. W. Pennebaker (Ed.). Emotion, disclosure, and health, (pp. 125-154). Washington, DC: American Psychological Association.

Schmitz, N., Hartkamp, N., Brinschwitz, C., & Michalek, S. (1999) Computerized administration of the Symptom Checklist (SCL-90-R) and the Inventory of Interpersonal Problems (IIP-C) in psychosomatic outpatients. Psychiatric Research, 87, 217-221.

Schulenberg, S.E., & Yutrzenka, B.A. (1999). The equivalence of computerized and paper-and-pencil psychological instruments: Implications for measures of negative affect. Behavior Research Methods, Instruments & Computers, 31, 315-321.

Schwartz, S.J., Mullis, R.L., & Dunham, R.M. (1998) Effects of authoritative structure in the measurement of identity formation: Individual computer-managed versus group paper-and-pencil testing. Computers in Human Behavior, 14, 239-248.

Stanton, J.M. (1998) An empirical assessment of data collection using the internet. Personnel Psychology, 51, 709-725.

Steiger, J.H. (1989). EzPATH: A supplementary module for SYSTAT and SYGRAPH. Evanstron, IL: SYSTAT, Inc.

Steiger, J.H. (1995). SEPATH: Structural equation modeling. In Statistica for Windows. Tulsa, OK: Statsoft, Inc.

Steiger, J.H. (1998). A note on multiple sample extensions of the RMSEA fit index. Structural Equation Modeling, 5, 411-419.

Steiger, J.H., & Fouladi, R.T. (1997). Noncentrality interval estimation and the evaluation of statistical models. In L Harlow, S Mulaik, and JH Steiger (Eds.) What if there were no significance tests. New Jersey: Lawrence Erlbaum.

Stone, A.A., & Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavioral medicine. Annals of Behavioral Medicine,16, 199-202.

Tseng, H., Macleod, H.A., & Wright, P. (1997). Computer anxiety and measurement of mood change. Computers in Human Behavior, 13, 305-316.

Wainer, H. (2000). Computerized adaptive testing. Mahwah, NJ: Lawrence Erlbaum.

Webster, J., & Compeau, D. (1996). Computer-assisted versus paper-and-pencil administration of questionnaires. Behavior Research Methods, Instruments and Computers, 28, 567-576.

Table 1.

Item statistics for paper-and-pencil (P&P) and online internet administrations of attachment and emotional function measures.

		P&P							Online						
	Items	N	M_{Item}	s^2_{Item}	r_{Item}	r_{Total}	α	95%CI: α	α_{Del}	N	M_{Item}	s^2_{Item}	r_{Item}	r_{Total}	α	95%CI: α	α_{Del}
IPPA Mother	25	156	3.93	1.27	.474	.673	.956	.945-.965	.956	183	4.10	1.05	.374	.590	.933	.918-.946	.934
IPPA Father	25	150	3.67	1.34	.459	.661	.954	.942-.964	.956	168	3.75	1.25	.464	.665	.955	.944-.964	.955
NMRS-C	10	161	3.42	1.08	.201	.379	.720	.651-.781	.721	213	3.43	.89	.285	.456	.802	.760-.839	.803
NMRS-B	10	163	3.54	1.15	.189	.360	.699	.626-.764	.703	220	3.55	.95	.245	.429	.763	.713-.807	.769
NMRS-G	10	161	3.57	1.01	.405	.593	.872	.841-.900	.874	221	3.53	.86	.448	.629	.888	.865-.909	.891
TMMS-A	13	162	3.88	.91	.349	.550	.873	.843-.900	.870	202	3.86	.74	.326	.526	.857	.826-.884	.860
MS-C	11	158	3.55	1.04	.402	.593	.877	.847-.904	.874	220	3.46	.83	.343	.536	.847	.815-.875	.847
TMMS-R	6	164	3.54	1.12	.399	.557	.801	.749-.864	.802	214	3.61	1.05	.405	.475	.804	.760-.842	.809

Note. M_{Item} = Mean item mean, s^2_{Item} =Mean item variance, r_{Item} = Mean inter-item correlation, r_{Total} = Mean corrected item-total correlation, α= Coefficient alpha (Average measure intraclass correlation (2-way mixed effect model, people random, items fixed, consistency definition); α_{Del} =Maximum coefficient alpha if single item deleted. NMRS-C = NMRS Cognitive Scale; NMRS-B = NMRS Behavior Scale; NMRS-G = NMRS General Scale; TMMS-A = TMMS Attention Scale; TMMS-C = TMMS Clarity Scale; TMMS-R = TMMS Repair Scale.

Table 2.

Summary results for multivariate analyses of the effect of mode of administration across the items on each scale as a function.

	N		Tests Of Covariance Matrices					Tests of Mean Vectors					
	P&P	O	Box's M	F	df_1	df_2	p	Wilks' Λ	F	df_1	df_2	p	ε^2
IPPA-M	156	183	618.04	1.76	325	327744.7	<.001	.858	2.07	25	313	.002	.142
IPPA-F	150	168	463.54	1.31	325	295431.5	<.001	.920	1.02	25	292	.444	.008
NMRS-C	161	213	100.33	1.77	55	384026.4	<.001	.939	2.34	10	363	.011	.061
NMRS-B	163	220	109.48	1.930	55	394453.5	<.001	.949	2.00	10	372	.032	.051
NMRS-G	161	221	89.17	1.58	55	385135.3	.004	.974	.991	10	371	.450	.026
TMMS-A	163	209	226.26	2.4	91	380426.7	<.001	.863	4.4	13	358	<.001	.137
TMMS-C	159	209	150.48	2.21	66	369577.0	<.001	.922	2.75	11	356	.002	.078
TMMS-R	161	218	42.96	2.01	21	436189.5	.004	.945	3.64	6	372	.002	.055

Note. P&P: Paper-and-pencil, O: Online. IPPA-M = IPPA – Mother Scale; IPPA-F = IPPA Father Scale; NMRS-C = NMRS Cognitive Scale; NMRS-B = NMRS Behavior Scale; NMRS-G = NMRS General Scale; TMMS-A = TMMS Attention Scale; TMMS-C = TMMS Clarity Scale; TMMS-R = TMMS Repair Scale.

Table 3

Summary results for multivariate general linear models across items in each attachment and emotional function instrument.

	df	Effect	Wilks' Λ	F	p	ε^2
IPPA Mother	25, 305	Sex	.886	1.57	.044	.114
		Ethnic Id	.940	.776	.772	.060
		Mode	.858	2.02	.003	.142
		Sex × Ethnic Id	.933	.879	.635	.067
		Mode × Sex	.972	.356	.998	.028
		Mode × Ethnic Id	.921	1.043	.410	.079
IPPA Father	25, 286	Sex	.889	1.44	.086	.111
		EthnicId	.956	.53	.970	.044
		Mode	.921	.98	.498	.079
		Sex × Ethnic Id	.907	1.18	.260	.093
		Mode × Sex	.899	1.28	.171	.101
		Mode × Ethnic Id	.950	.60	.936	.050
NMRS	30, 318	Sex	.845	1.95	.003	.155
		Ethnic Id	.877	1.49	.052	.123
		Mode	.911	1.04	.419	.089
		Sex × Ethnic Id	.891	1.30	.140	.109
		Mode × Sex	.923	.89	.644	.077
		Mode × Ethnic Id	.928	.83	.727	.072
TMMS	30, 309	Sex	.851	1.80	.008	.149
		Ethnic Id	.886	1.33	.122	.114

Mode	.891	1.26	.170	.109
Sex × Ethnic Id	.884	1.35	.112	.116
Mode × Sex	.921	.88	.649	.079
Mode × Ethnic Id	.905	1.08	.364	.095

Table 4.

Goodness of fit measures: RMSEA and 90% confidence interval lower and upper bounds for each measure.

		df	ML				SB			
			χ^2	RMSE-A	LB	UB	χ^2	RMSE-A	LB	UB
IPPA-M	MU	548	**1890.7**	.116	.109	.120				
	MC1	573	**1901.5**	.112	.106	.117				
	MC2	599	**1766.6**	.126	.117	.132				
	D1	25	9.8	.000	.000	.045				
	D2	26	0.0	.000	.000	.044				
	O	274	**1174.7**	.115	.108	.121	**872.9**	.094	.087	.099
	P	274	**859.2**	.128	.118	.137	**664.6**	.104	.095	.113
	I	274	**786.9**	.126	.115	.136	**582.6**	.098	.087	.107
IPPA-F	MU	548	**1531.7**	.120	.113	.127				
	MC1	573	**1562.6**	.117	.110	.124				
	MC2	599	**1601.1**	.116	.109	.122				
	D1	25	30.9	.031	.000	.058				
	D2	26	38.5	.044	.000	.050				
	O	274	**1125.6**	.112	.105	.118	**887.9**	.095	.089	.101
	P	274	**766.6**	.112	.117	.107	**635.2**	.100	.100	.090
	I	274	**765.1**	.124	.112	.133	**592.1**	.099	.088	.109
NMRS	MU	798	**1551.4**	.088	.081	.093				
	MC1	831	**1601.9**	.086	.079	.092				
	MC2	864	**1700.0**	.089	.082	.093				

	D1	33	50.5	.046	.000	.068				
	D2	33	98.9	.089	.068	.109				
	O	399	**1037.4**	.080	.074	.086	**805.0**	.064	.058	.070
	P	399	**765.9**	.084	.075	.092	**680.1**	.073	.064	.082
	I	399	**785.5**	.091	.081	.099	**532.5**	.053	.041	.063
TMMS	MU	798	**2776.3**	.141	.134	.146				
	MC1	831	**2864.0**	.140	.134	.146				
	MC2	864	**2910.9**	.139	.132	.143				
	D1	33	87.7	.081	.059	.101				
	D2	33	46.9	.041	.000	.064				
	O	399	**2126.4**	.132	.126	.137	**1027.2**	.079	.075	.084
	P	399	**1392.2**	.134	.129	.145	**592.2**	.061	.052	.068
	I	399	**1384.1**	.145	.136	.152	205.6	.000	.000	.000

Note. Bold χ^2 values are statistically significant at $p < .05$. ML=Maximum likelihood, B=Satorra-Bentler scaled maximum likelihood; RMSEA=Root mean square error of approximation, LB=90%CI lower bound, UB=90%CI upper bound, MU=Multisample unconstrained, MC1=Multisample factor loadings and intercorrelations constrained, MC2=Multisample factor loadings, intercorrelations, variances, and unique variances constrained, D1= Difference test of MU with MC1, D2= Difference test of MC1 with MC2c O=Single sample overall, P=Single sample paper-and-pencil study, I=Single sample online internet study.

Table 5.

Descriptive statistics for pooled data set scale scores on instruments of attachment and emotional function.

	N	Min[a]	Max[a]	M	SD	Skew[b]	Kurt.[b]	r[c]							
								IPPA Mother	IPPA Father	NMRS Cognit.	NMRS Behav.	NMRS General	TMMS Atten.	TMMS Clarity	TMMS Repair
IPPA Mother	379	31	125	99.79	18.37	-1.18	1.33	1.000							
IPPA Father	359	29	125	91.65	20.45	-.77	.09	.369	1.000						
NMRS Cognitive	390	1.80	4.90	3.43	.56	-.37	.14	.200	.221	1.000					
NMRS Behavior	391	1.80	5.00	3.55	.55	-.50	.37	.295	.265	.556	1.000				
NMRS General	393	1.10	5.00	3.53	.67	-.54	.41	.230	.258	.749	.600	1.000			
TMMS Attention	384	1.77	4.85	3.72	.54	-.29	.21	.241	.144	.169	.311	.209	1.000		
TMMS Clarity	391	1.30	4.91	3.61	.58	-.43	.46	.182	.221	.348	.360	.485	.317	1.000	
TMMS Repair	372	1.33	5.00	3.66	.66	-.52	.19	.258	.247	.529	.504	.618	.349	.500	1.000

Note. a: Possible range for IPPA Mother & IPPA Father is 25-125; all others 1-5. b: SE skew ≈.12, SE kurtosis ≈.25. c: All p's<.001.